



PATENT OFFICE
JAPANESE GOVERNMENT

This is to certify that the annexed is a true copy of the following application as filed with this Office.

Date of Application: February 16, 1999
Application Number : P11-037525
Applicant(s) : KDD CORPORATION

December 10, 1999

Commissioner,
Patent Office Takahiko KONDOU

H11-3085578

日 本 国 特 許 庁

PATENT OFFICE
JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日

Date of Application:

1999年 2月16日

出 願 番 号

Application Number:

平成11年特許願第037525号

出 願 人

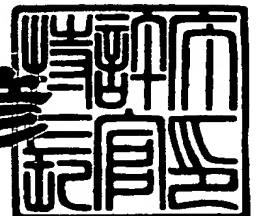
Applicant(s):

ケイディディ株式会社

1999年12月10日

特許庁長官
Commissioner,
Patent Office

近 藤 隆 彦



出証番号 出証特平11-3085578

【書類名】 特許願

【整理番号】 KDD-30

【提出日】 平成11年 2月16日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/00

【発明の名称】 URL階層構造を利用した情報自動フィルタリング方法
および装置

【請求項の数】 4

【発明者】

 【住所又は居所】 埼玉県上福岡市大原 2－1－1 5 株式会社ケイディデ
 ィ研究所内

 【氏名】 帆足 啓一郎

【発明者】

 【住所又は居所】 埼玉県上福岡市大原 2－1－1 5 株式会社ケイディデ
 ィ研究所内

 【氏名】 井ノ上 直己

【発明者】

 【住所又は居所】 埼玉県上福岡市大原 2－1－1 5 株式会社ケイディデ
 ィ研究所内

 【氏名】 橋本 和夫

【特許出願人】

 【識別番号】 000001214

 【氏名又は名称】 ケイディディ株式会社

【代理人】

 【識別番号】 100083806

 【弁理士】

 【氏名又は名称】 三好 秀和

 【電話番号】 03-3504-3075

【選任した代理人】

【識別番号】 100100712

【弁理士】

【氏名又は名称】 岩▲崎▼ 幸邦

【選任した代理人】

【識別番号】 100095500

【弁理士】

【氏名又は名称】 伊藤 正和

【選任した代理人】

【識別番号】 100101247

【弁理士】

【氏名又は名称】 高橋 俊一

【手数料の表示】

【予納台帳番号】 001982

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9506777

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 URL階層構造を利用した情報自動フィルタリング方法および装置

【特許請求の範囲】

【請求項 1】 インターネットを介して提供される各種情報のうち不適切情報を識別し、この識別した不適切情報の提供を阻止する情報自動フィルタリング方法であって、

インターネットを介して提供されるHTML情報を入力し、このHTML情報のURLが上位URLであるか否かを判定し、

このURLが上位URLである場合、この上位URLが示す情報に出現する単語を抽出し、この抽出された単語に基づいて該情報が不適切であるか否かの自動フィルタリングを行い、

この自動フィルタリングの結果、前記情報が不適切であると判定された場合、前記上位URLを不適切上位URL一覧に登録するとともに、前記情報の提供を阻止し、

前記HTML情報のURLが上位URLでなかった場合、このURLを前記登録された不適切上位URL一覧の各URLと照合して、一致するURLがあるか否かを判定し、一致する場合、このURLが示す情報の提示を阻止し、

前記URLが不適切上位URL一覧のURLと一致するものがない場合、該URLが示す情報に出現する単語を抽出し、この抽出された単語に基づいて該情報が不適切であるか否かの自動フィルタリングを行い、

この自動フィルタリングの結果、前記情報が不適切であると判定された場合、該情報の提供を阻止すること

を特徴とするURL階層構造を利用した情報自動フィルタリング方法。

【請求項 2】 不適切な情報を提供するURLを不適切URL一覧として登録しておき、前記入力されたHTML情報のURLを前記不適切URL一覧の各URLと照合して、一致するURLがあるか否かを判定し、一致する場合、このURLが示す情報の提示を阻止する第三者判定に基づくフィルタリングを更に行うことを特徴とする請求項 1 記載のURL階層構造を利用した情報自動フィルタ

リング方法。

【請求項3】 インターネットを介して提供される各種情報のうち不適切情報を識別し、この識別した不適切情報の提供を阻止する情報自動フィルタリング装置であって、

インターネットを介して提供されるHTML情報を入力する入力手段と、

この入力されたHTML情報のURLが上位URLであるか否かを判定する上位URL判定手段と、

該上位URL判定手段による判定の結果、前記URLが上位URLである場合、この上位URLが示す情報に出現する単語を抽出し、この抽出された単語に基づいて該情報が不適切であるか否かの自動フィルタリングを行う第1の自動フィルタリング手段と、

この自動フィルタリングの結果、前記情報が不適切であると判定された場合、該情報の提示を阻止するとともに、前記上位URLを不適切上位URL一覧テーブルに登録する不適切上位URL一覧登録手段と、

前記上位URL判定手段による判定の結果、前記HTML情報のURLが上位URLでなかった場合、このURLを前記不適切上位URL一覧テーブルに登録されている各URLと照合して、一致するURLがあるか否かを判定する不適切URL判定手段と、

この判定の結果、前記URLが不適切上位URL一覧テーブルに登録されているURLと一致するものがない場合、該URLが示す情報に出現する単語を抽出し、この抽出された単語に基づいて該情報が不適切であるか否かの自動フィルタリングを行う第2の自動フィルタリング手段と、

前記不適切URL判定手段による判定の結果、前記URLが不適切上位URL一覧テーブルに登録されているURLと一致する場合、このURLが示す情報の提示を阻止し、また前記第2の自動フィルタリング手段によるフィルタリングの結果、前記情報が不適切であると判定された場合、該情報の提供を阻止する情報提示阻止手段と

を有することを特徴とするURL階層構造を利用した情報自動フィルタリング装置。

【請求項4】 不適切な情報を提供するURLを不適切URL一覧テーブルに登録する不適切URL一覧登録手段と、

前記入力手段から入力される前記HTML情報のURLを前記不適切URL一覧テーブルに登録されている各URLと照合し、一致するURLがあるか否かを判定する一致URL判定手段と、

この判定の結果、一致するURLがある場合、このURLが示す情報の提示を阻止する第三者判定に基づくフィルタリング手段と

を更に有することを特徴とする請求項3記載のURL階層構造を利用した情報自動フィルタリング装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、インターネットを介して提供される各種情報のうち不適切情報、例えばポルノ画像等のような有害情報を識別し、この識別した不適切情報の提供を阻止する有害情報自動フィルタリング方法および装置に関し、更に詳しくは、階層構造に構成されているURLに基づき不適切情報を判定し、その不適切情報の提示を阻止するURL階層構造を利用した情報自動フィルタリング方法および装置に関する。

【0002】

【従来の技術】

インターネットの急速な広がりに伴い、限られた専門家の道具でしかなかったコンピュータはごく一般の家庭や学校などにも導入され始めている。このため、これまでコンピュータに触れることすらなかった多くの一般人でも気軽にインターネットにアクセスすることが可能になった。こうした背景の中、近年深刻な問題となっているのがインターネット上に氾濫するポルノ画像などの有害情報に対する子供のアクセスである。この問題に対処するため、アメリカでは政府機関がインターネット上の情報を検閲することを可能にした「通信品位法」という法律が提案されたが、裁判の結果、表現の自由を保証する憲法に違反すると判決され、立法することができなかった。

【0003】

そこで最近注目されているのが「情報フィルタリング」という技術である。情報フィルタリングとは、ユーザがインターネット上の情報にアクセスする際にその情報の有害性をチェックし、有害と判定された場合は何らかの手段によりその情報へのアクセスをブロックするという技術である。

【0004】

現在市販されている有害情報フィルタリングソフトで取り入れられている手法は大きく以下の3つに分類される。

【0005】

- (1) 自己判定によるフィルタリング
- (2) 第三者の判定によるフィルタリング
- (3) 自動フィルタリング

ここではこの3つの手法について簡単に解説する。まず、自己判定によるフィルタリング手法ではWWW情報の提供者が自らのコンテンツの有害性について判定を行い、その結果をHTMLファイル内に記述する。フィルタリングソフトはこの記述された結果を参照し、有害と判断された場合にアクセスをブロックする。この手法によるフィルタリングを図6に示す。

【0006】

図6に示す自己判定に基づくフィルタリングでは、米国マサチューセッツ工科大学のWorld Wide Web Consortium が作成したPICS (Platform for Internet Content Selection) と呼ばれるインターネットコンテンツの評価を記述するための基準を使用している。PICSを使用することにより、コンテンツ提供者は簡単に自分の提供している情報を描写し、開示することができる。

【0007】

多くの場合、コンテンツ提供者がこのような評価結果を公開する際には、PICSによる評価結果を出力する評価機関のサービスを利用する。このような評価機関の代表として、Recreational Software Advisory Council (RSAC) やSafeSurfといった団体があげられ、それぞれ独自に設定した基準による評価結果を提供している。コンテンツ提供者はこれらの機関からの評価結果をHTMLファ

イルのヘッダに記述する。図7にこの評価結果の記述例を示す。

【0008】

この自己判定はコンテンツ提供者の自主性に任せられるというのが現状である。そのため、多くのコンテンツ提供者がこの判定を受けようという意志を持たない限りは本手法による有効な有害情報フィルタリングは不可能であるといえる。

【0009】

次に、第三者による判定に基づくフィルタリングについて説明する。有害情報フィルタリングソフトを作成している業者の中には、WWW上のホームページの有害性を独自に判定し、その結果をフィルタリングソフトの判断基準とする手法を取り入れている。一般的には、この評価の結果として有害なホームページのURL一覧が構築される。このURLのリストはフィルタリングソフトとともにユーザに分配され、フィルタリングソフトの判断基準となる。多くの場合、フィルタリングソフトはこの有害URL一覧を定期的にダウンロードする仕組みになっている。第三者による判定に基づく有害情報フィルタリングの仕組みを図8に示す。

【0010】

このような仕組みを持つソフトウェアの代表的なものとしてCyberPatrol があげられる。CyberPatrol は「暴力」「性行為」など13個のジャンルに対し、それぞれ有害URL一覧を持っており、これらのシステムに従って有害情報フィルタリングを行う。

【0011】

この手法で使用される有害URL一覧はそれぞれのソフトウェア業者でホームページをアクセスし、判定を行うことによって作成・拡張されているため、新しく設立されたホームページや従来のURLから別のURLに移動したホームページには対処することは不可能である。従って、こうした評価対象外のページに対するフィルタリングには対処できないのが現状である。

【0012】

次に、自動フィルタリングについて説明する。有害情報フィルタリングソフトの中にはアクセスされたホームページの中身をチェックし、有害性の判断を行う

ものもある。このような発想は初期のフィルタリングソフトで導入されていた。その例として、例えば“sex”や“xxx”といった文字列がURLに含まれていた場合、そのURLへのアクセスを禁止するなどという処理を行うソフトが存在した。現在はページの中身について検証を行うソフトも開発されている。CyberSITTER はこうした自動フィルタリングを行うソフトの1つである。このソフトではアクセスされたページに含まれる有害な単語を取り除いて出力するという手法によってフィルタリングが行われる。

【0013】

本手法には2つの問題点がある。まず1つは、この自動判定を行う際に生じる処理時間である。最も、この程度の処理では数ミリ秒程度の少ない処理時間ではあるが、こうした短い時間でもユーザにフラストレーションが生じる可能性は否定できない。

【0014】

もう一方の問題は、自動フィルタリングの精度である。まず、単語単位で有害性を判断するような判定アルゴリズムが採用されている場合、多くの無害なページがブロックされてしまう可能性が高い。現に、イギリスの“Sussex”という町に関するホームページがブロックされるといった悪例も報告されている。更に、ページ内のテキスト情報のみに着目して自動フィルタリングを行う場合、画像のみが表示されているページをブロックすることは不可能であるという問題もあげられる。

【0015】

【発明が解決しようとする課題】

フィルタリングソフトの大きな目的は有害なページがブロックされる割合を増やすことと、無害なページが誤ってブロックされる割合を減らすことである。ブロックされたページのうち、実際に有害だったページの割合を正解率（precision）、実際に有害なページのうちブロックされたページの割合を再現率（recall）とすると、フィルタリングソフトの目的は正解率と再現率をともに高めることであるといえる。

【0 0 1 6】

上述した各手法にはそれぞれ一長一短がある。各手法の特徴を正解率と再現率という観点からまとめた。この結果を表 1 に示す。

【0 0 1 7】

【表 1】

表 1 各フィルタリング手法とその正解率および再現率

手 法	正 解 率	再 現 率
自己判定	低	低
第三者判定	高	低
自動フィルタリング	低	高

このように、現在市販されているフィルタリングソフトでは十分なフィルタリング性能が得られないのが現状である。

【0 0 1 8】

また、上述したように、従来の自動フィルタリングでは、ページ内のテキスト情報のみに着目して自動フィルタリングを行うため、テキスト情報が少ないかまたは全くなく、画像のみが表示されているページを阻止することが不可能であるという問題がある。

【0 0 1 9】

本発明は、上記に鑑みてなされたもので、その目的とするところは、階層構造になっている URL の上位 URL を用いることにより正解率および再現率共に向上し得るとともに、画像のみが掲載されている少テキストページに対しても内容の不適切さを適確に判定し得る URL 階層構造を利用した情報自動フィルタリング方法および装置を提供することにある。

【0 0 2 0】

【課題を解決するための手段】

上記目的を達成するため、請求項 1 記載の本発明は、インターネットを介して提供される各種情報のうち不適切情報を識別し、この識別した不適切情報の提供を阻止する情報自動フィルタリング方法であって、インターネットを介して提供

されるHTML情報を入力し、このHTML情報のURLが上位URLであるか否かを判定し、このURLが上位URLである場合、この上位URLが示す情報に出現する単語を抽出し、この抽出された単語に基づいて該情報が不適切であるか否かの自動フィルタリングを行い、この自動フィルタリングの結果、前記情報が不適切であると判定された場合、前記上位URLを不適切上位URL一覧に登録するとともに、前記情報の提供を阻止し、前記HTML情報のURLが上位URLでなかった場合、このURLを前記登録された不適切上位URL一覧の各URLと照合して、一致するURLがあるか否かを判定し、一致する場合、このURLが示す情報の提示を阻止し、前記URLが不適切上位URL一覧のURLと一致するものがない場合、該URLが示す情報に出現する単語を抽出し、この抽出された単語に基づいて該情報が不適切であるか否かの自動フィルタリングを行い、この自動フィルタリングの結果、前記情報が不適切であると判定された場合、該情報の提供を阻止することを要旨とする。

【 0 0 2 1 】

請求項 1 記載の本発明にあっては、入力されたHTML情報のURLが上位URLである場合、この上位URLが示す情報に対して自動フィルタリングを行い、その結果、該情報が不適切である場合、前記上位URLを不適切上位URL一覧に登録するとともに、前記情報の提供を阻止し、上位URLでなかった場合、このURLを不適切上位URL一覧の各URLと照合し、一致するURLがある場合、このURLが示す情報の提示を阻止し、一致するものがない場合、該URLが示す情報に対して自動フィルタリングを行い、その結果、前記情報が不適切である場合、該情報の提供を阻止するため、画像のみが提示されている少テキストページでもその不適切さを適確に判定して阻止することができ、正解率および再現率の両方を向上することができる。

【 0 0 2 2 】

また、請求項 2 記載の本発明は、請求項 1 記載の発明において、不適切な情報を提供するURLを不適切URL一覧として登録しておき、前記入力されたHTML情報のURLを前記不適切URL一覧の各URLと照合して、一致するURLがあるか否かを判定し、一致する場合、このURLが示す情報の提示を阻止す

る第三者判定に基づくフィルタリングを更に行うことを要旨とする。

【0023】

請求項2記載の本発明にあっては、不適切な情報を提供するURLを不適切URL一覧として登録しておき、HTML情報のURLを不適切URL一覧の各URLと照合して、一致するURLがある場合、このURLが示す情報の提示を阻止する第三者判定に基づくフィルタリングを更に行うため、この第三者判定に基づくフィルタリングと上位URLを利用した自動フィルタリングの両方により更に完全にフィルタリングを行うことができる。

【0024】

更に、請求項3記載の本発明は、インターネットを介して提供される各種情報のうち不適切情報を識別し、この識別した不適切情報の提供を阻止する情報自動フィルタリング装置であって、インターネットを介して提供されるHTML情報を入力する入力手段と、この入力されたHTML情報のURLが上位URLであるか否かを判定する上位URL判定手段と、該上位URL判定手段による判定の結果、前記URLが上位URLである場合、この上位URLが示す情報に出現する単語を抽出し、この抽出された単語に基づいて該情報が不適切であるか否かの自動フィルタリングを行う第1の自動フィルタリング手段と、この自動フィルタリングの結果、前記情報が不適切であると判定された場合、該情報の提示を阻止するとともに、前記上位URLを不適切上位URL一覧テーブルに登録する不適切上位URL一覧登録手段と、前記上位URL判定手段による判定の結果、前記HTML情報のURLが上位URLでなかった場合、このURLを前記不適切上位URL一覧テーブルに登録されている各URLと照合して、一致するURLがあるか否かを判定する不適切URL判定手段と、この判定の結果、前記URLが不適切上位URL一覧テーブルに登録されているURLと一致するものがない場合、該URLが示す情報に出現する単語を抽出し、この抽出された単語に基づいて該情報が不適切であるか否かの自動フィルタリングを行う第2の自動フィルタリング手段と、前記不適切URL判定手段による判定の結果、前記URLが不適切上位URL一覧テーブルに登録されているURLと一致する場合、このURLが示す情報の提示を阻止し、また前記第2の自動フィルタリング手段によるフィ

ルタリングの結果、前記情報が不適切であると判定された場合、該情報の提供を阻止する情報提示阻止手段とを有することを要旨とする。

【0025】

請求項3記載の本発明にあっては、入力されたHTML情報のURLが上位URLである場合、この上位URLが示す情報に対して自動フィルタリングを行い、その結果、該情報が不適切である場合、前記上位URLを不適切上位URL一覧テーブルに登録するとともに、前記情報の提供を阻止し、上位URLでなかった場合、このURLを不適切上位URL一覧テーブルの各URLと照合し、一致するURLがある場合、このURLが示す情報の提示を阻止し、一致するものがない場合、該URLが示す情報に対して自動フィルタリングを行い、その結果、前記情報が不適切である場合、該情報の提供を阻止するため、画像のみが提示されている少テキストページでもその不適切さを適確に判定して阻止することができる。正解率および再現率の両方を向上することができる。

【0026】

請求項4記載の本発明は、請求項3記載の発明において、不適切な情報を提供するURLを不適切URL一覧テーブルに登録する不適切URL一覧登録手段と、前記入力手段から入力される前記HTML情報のURLを前記不適切URL一覧テーブルに登録されている各URLと照合し、一致するURLがあるか否かを判定する一致URL判定手段と、この判定の結果、一致するURLがある場合、このURLが示す情報の提示を阻止する第三者判定に基づくフィルタリング手段とを更に有することを要旨とする。

【0027】

請求項4記載の本発明にあっては、不適切な情報を提供するURLを不適切URL一覧テーブルとして登録しておき、HTML情報のURLを不適切URL一覧テーブルの各URLと照合して、一致するURLがある場合、このURLが示す情報の提示を阻止する第三者判定に基づくフィルタリングを更に行うため、この第三者判定に基づくフィルタリングと上位URLを利用した自動フィルタリングの両方により更に完全にフィルタリングを行うことができる。

【0 0 2 8】

【発明の実施の形態】

以下、図面を用いて本発明の実施の形態について説明する。図 1 は、本発明の一実施形態に係る URL 階層構造を利用した情報自動フィルタリング装置の構成を示すブロック図である。

【0 0 2 9】

図 1 に示す情報自動フィルタリング装置は、インターネットを介して提供される各種情報のうち不適切情報、例えばポルノ等の有害情報を識別し、この識別した不適切情報の提供を阻止するものであり、インターネットを介して提供される HTML 情報を入力する入力部 1、この入力された情報に出現する単語を抽出する単語抽出部 3、この抽出した単語や本実施形態の情報自動フィルタリング処理を実行するソフトウェアやその他の各種情報を記憶する記憶部 5、単語の重みデータを格納する単語重みデータ格納部 7、自動フィルタリングを行う自動フィルタリング部 9、有害上位ページ一覧をテーブルとして格納している有害上位ページ一覧テーブル格納部 11、およびフィルタリングした結果を出力する出力部 13 から構成されている。

【0 0 3 0】

本実施形態の情報自動フィルタリング装置は、URL 階層構造を有する URL のうち上位 URL を利用して、有害情報のフィルタリングを行うものであるが、まずその概念について説明する。

【0 0 3 1】

上述したように、自動フィルタリングの大きな問題の 1 つとしてテキスト情報が少ないあるいは全くないホームページに対するフィルタリングが困難であることがあげられる。特にポルノ系の有害情報ページには画像のみが掲載されているものが多数含まれていると考えられるため、これらの少テキストページに対する対処法を検討する必要がある。しかし、典型的な WWW ユーザならば、画像のみのページにアクセスするためにリンクをたどるものと考えられる。この仮定が真実ならば、画像ページに至るまでの上位階層のページに対してフィルタリングを行えば画像ページへのアクセスもブロックすることが可能になる。この上位階層

のページに対してフィルタリングを行う手法について説明する。

【0032】

まず、上位階層にあるページをそのURLが以下に列挙した7つの文字列で終わるページとする：

- (1) index.html
- (2) index.htm
- (3) index.shtml
- (4) welcome.html
- (5) welcome.htm
- (6) welcome.shtml
- (7) /

例えば、<http://www.kdd.co.jp/index.html> や <http://www.asahi.com/> などは上位ページとみなされる。これら上位ページのうち、フィルタリングソフトによって有害と判断されたものを有害上位ページ一覧に保存する。この際、URLのすべてを保存するのではなく、URLのうち最も深いディレクトリまでのURLを保存することにする。例えば、<http://www.###.co.jp/index.html> が有害の場合は、<http://www.###.co.jp/> を一覧に保存し、<http://www.###.co.jp/aaa/bbb/ccc/index.html> が有害の場合は<http://www.###.co.jp/aaa/bbb/ccc/> を一覧に保存する。

【0033】

上位ページ以外のページへのアクセスの際、通常の自動有害性判断の前に、この有害上位ページ一覧に記されているURLとアクセスされているページのURLを比較する。比較の結果、アクセスされたページのURLのディレクトリが有害上位ページ一覧中のどれかのURLと一致した場合、そのページを有害であるとみなす。例えば<http://www.###.co.jp/> が有害上位ページ一覧に含まれていた場合、<http://www.###.co.jp/aaa/bbb.html> も<http://www.###.co.jp/nantoka.html> も有害であるとみなす。一方、有害上位ページ一覧中のデータと一致しない場合は、自動フィルタリングソフトにより有害性の判断を行う。

【0034】

上述した考え方に基づいて本実施形態の情報自動フィルタリング装置は有害情報を阻止するように構成されている。次に、図2に示すフローチャートを参照して、図1に示すURL階層構造を利用した情報自動フィルタリング装置の作用について説明する。

【0035】

図2において、まず入力部1からインターネットを介して提供されるHTML情報であるHTML文書が入力されると（ステップS11）、この入力されたHTML文書のURLが上位URL、すなわち上位ページであるか否かが判定される（ステップS13）。この判定の結果、前記HTML文書のURLが上位URLである場合には、この上位URLが示す文書、すなわち情報に出現する単語を単語抽出部3で抽出し、この抽出した単語に基づいて自動フィルタリング部9による自動フィルタリングを行い（ステップS15）、前記情報が有害であるか否かについての判定を行う（ステップS17）。

【0036】

この自動フィルタリングの情報の有害性について判定の結果、前記情報が有害である場合には、前記上位URLを有害上位ページ一覧テーブル格納部11の有害上位ページ一覧テーブルに登録するとともに（ステップS21）、この情報の提供を阻止（ブロック）して処理を終了する（ステップS31）。

【0037】

一方、ステップS17における判定の結果、前記情報が有害でない場合には、出力部13によりブラウザに表示して処理を終了する（ステップS19）。

【0038】

また、ステップS13における上位ページか否かの判定の結果、上位ページでない場合には、このURLを有害上位ページ一覧テーブル格納部11に有害上位ページ一覧テーブルとして登録されている各URLと照合し（ステップS23）、一致するURLがあるか否かをチェックする（ステップS25）。このチェックの結果、有害上位ページ一覧テーブルに一致するURLがある場合には、このURLが示す情報の提供を阻止して処理を終了する（ステップS31）。

【0039】

ステップS25における一致するURLがあるか否かのチェックの結果、一致するURLがない場合には、このURLが示す情報に出現する単語を単語抽出部3で抽出し、この抽出した単語に基づいて自動フィルタリング部9による自動フィルタリングを行い（ステップS27）、前記情報が有害であるか否かについての判定を行う（ステップS29）。

【0040】

この自動フィルタリングの情報の有害性について判定の結果、前記情報が有害である場合には、該情報の提供を阻止して処理を終了するが（ステップS31）、ステップS29における判定の結果、前記情報が有害でない場合には、出力部13によりブラウザに表示して処理を終了する（ステップS19）。

【0041】

上述した本実施形態のURL階層構造を利用した情報自動フィルタリング装置の効果を判定するため、次に示すような評価実験を行った。

【0042】

この評価実験における評価用のデータとして、WWW上のデータを自動的に収集するソフト（「収集ロボット」）を使用して大量の有害ページを収集した。この収集ロボットは290個の有害ページ（うち、160個が日本語、130個が英語で記述されたページ）へのリンクが張られているHTMLページから起動され、順々にリンクをたどりその途中でアクセスされたHTML文書を収集した。なお、この際収集されたデータはHTML文書のみであり、画像データ、音声データ等は収集していない。この結果、28034個のHTML文書が収集された。

【0043】

次に、収集された個々のHTML文書に対し、主観評価によってその有害性を3段階で評価した。なお、この評価は各ページの性的表現の有無についてのものである。各評価段階の基準を表2に示す。

【0044】

【表2】

表2 HTMLページの有害性評価基準

レベル	評価	説明
1	Safe (無害)	性的表現が全くないページ。
2	Partly Hazardous (やや有害)	直接的な性的表現はないものの、性的表現を暗に示す内容が記述されたページ。
3	Hazardous (有害)	性的表現が含まれているページ。

収集されたデータに対する有害性評価の結果は表3に示す通りである。

【0045】

【表3】

表3 HTMLデータの有害性評価結果

レベル	文書数	割合
1	6230	22.2%
2	2055	7.3%
3	19749	70.4%

この有害性評価とともに、各ページに記述されている言語についての調査も行った。その結果を表4に示す。

【0046】

【表4】

表4 HTMLデータの記述言語調査結果

言語	文書数	割合
英語	15448	55.1%
日本語	10839	38.7%
その他	1747	6.2%

また、このデータのうち、上述した「有害上位ページ」に該当するデータの全有害データ中の割合についても調査した。なお、ここでは有害データを前記有害性評価の結果がレベル2または3だったものとする。その結果を表5に示す。

【0047】

【表5】

表5 HTMLデータの上位・下位ページの有害データ中の割合

種 類	文 書 数	割 合
上位データ	1495	6.9%
下位データ	20309	93.1%

次に、図1に示した実施形態の情報自動フィルタリング装置に使用されている自動フィルタリング部9による自動フィルタリングのアルゴリズム、特に前記評価実験に使用した自動フィルタリングのアルゴリズムについて説明する。なお、この自動フィルタリングは、情報検索や自動分類等に使用されているベクトル空間モデルを使用している。

【0048】

まず、入力部1から入力されたHTML文書をベクトル空間モデルによって表現する。すなわち、すべての文書を表現するn個の単語を選択し、それぞれの文書をn次元のベクトルで次式のように表現する。

【0049】

【数1】

$$\vec{V}_d = (f_{d1}, \dots, f_{dn}) \quad \dots (1)$$

このベクトルの各要素は、各々単語の文書dでの出現頻度を正規化したものである。単語の出現頻度の正規化には次に示す数式で表されるTF*IDFという手法を用いている。

【0050】

【数2】

$$f_{di} = tf_{di} * \log \left(\frac{N}{df_i} \right) \quad \dots (2)$$

ここで、 tf_{di} は単語iが文書dに出現する頻度、Nはすべての文書の数、 df_i は単語iが出現する文書の数である。

【0051】

自動フィルタリングは、次に示す数式で表される線形識別関数によって行われ、この関数によって単語重みの総和 $Dis(d)$ が計算される。

【0052】

【数3】

$$Dis(d) = \sum_{i=1}^n w_i * f_{di} \quad \dots (3)$$

ここで、 w_i は各単語 i に対する重みであり、 f_{di} は上式 (3) の値であり、文書における各単語の f_{di} 値である。

【0053】

上述した式 (3) から、総和 $Dis(d)$ が 0 より大きい場合、前記文書は有害であり、0 以下である場合、無害であると判定される。

【0054】

なお、上述した各単語 i に対する重みは文書 d が有害な場合、総和 $Dis(d) > 0$ となり、無害な場合、総和 $Dis(d) \leq 0$ となるように設定される。

【0055】

次に、この単語の重みの設定について図3に示すフローチャートを参照して説明する。なお、この単語の重みの学習には **perceptron learning algorithm (PLA)** を使用している。

【0056】

図3においては、まず各種パラメータを設定する (ステップ S51)。このパラメータとしては、各単語の重みの集合 $W = (w_1, \dots, w_n)$ 、 N 個の学習データ $E = \{d_1, \dots, d_N\}$ 、定数 η 、最大学習回数 Max 、図3に示す学習処理を繰り返し行う学習回数 m がある。

【0057】

次に、単語の重みの集合 W を初期化する (ステップ S53)。この初期化では、各単語の重みに乱数を入力する。それから、すべての学習データに対して前記単語重みの総和 $Dis(d)$ を上式 (3) により計算する (ステップ S55)。

【0058】

そして、この計算の結果、すべての無害な文書 d について総和 $Dis(d) \leq 0$ であり、かつすべての有害な文書 d について総和 $Dis(d) > 0$ であるか否かをチェックし（ステップ S57）、そうである場合には、処理を終了するが、そうでない場合には、このように誤って分類されたすべての文書 d について次のステップ S61, S63 で示すように重みの変化度合 S を補正する（ステップ S59）。

【0059】

すなわち、ステップ S61 では、文書 d_i が有害であって、かつ総和 $Dis(d) \leq 0$ の場合には、重み変化度合 S を増加するように補正し、またステップ S63 では、文書 d_i が無害であって、かつ総和 $Dis(d) > 0$ の場合には、重み変化度合 S を低減するように補正する。

【0060】

そして、このように補正された重み変化度合 S を使用して単語重みの集合 W をステップ S65 で示す式のように補正する。それから、学習回数 m を +1 インクリメントし（ステップ S67）、この学習回数 m が最大学習回数 Max より小さいか否かをチェックし（ステップ S69）、また最大学習回数 Max より小さい場合には、ステップ S55 に戻り、ステップ S57 に示した条件が満たされるまで、ステップ S55 以降の処理を繰り返し行う。

【0061】

次に、上述した実施形態の URL 階層構造を利用した情報自動フィルタリング装置の評価実験について説明する。この評価実験は次に示す 3 つのプロセスからなる。

【0062】

- (1) 文書を表現する単語集合抽出。
- (2) 各単語に対する重みの学習。
- (3) 最終評価。

【0063】

まず、単語抽出のプロセスでは、収集されたデータの中から日本語で記述され

た文書 5912 個に対し形態素解析を行い、名詞・固有名詞・未定義語を抽出した。日本語用の形態素解析ソフトを使用したため、文書中に含まれる英単語は未定義語として抽出される。また、この形態素解析の際には標準の日本語辞書とともに辞書に載っていない性的表現などに関する用語集を制作し、これを使用した。この専門用語集には約 1000 語の単語が登録されている。また、抽出された単語のうち、データ全体での出現頻度が 20 以下の単語は取り除かれた。この結果、8013 個の単語が抽出された。

【0064】

重み学習では評価データの一部が使用された。この学習用データは 18387 個の HTML 文書から構成される。このうち、英語で記述された文書は 9263 個、日本語で記述された文書は 8171 個、その他の言語で記述された文書は 953 個であった。最終評価は単語抽出用データと学習データを含む評価データ全体に対して行われた。

【0065】

評価結果では、テキスト情報が少ない HTML 文書に対するフィルタリングが困難であるという仮定を証明するため、1 つの HTML 文書に出現する全単語数が閾値 min 以下の文書に対してフィルタリングを行い、その正解率と再現率を求めた。表 6 にその結果を示す。

【0066】

【表 6】

表 6 少テキスト HTML データに対する
自動フィルタリングの正解率と再現率

min	正 解 率	再 現 率
5	88.2%	6.5%
10	92.1%	41.4%
15	91.0%	44.8%
20	91.6%	49.5%

この結果から明らかなように、単語数が減るにつれ、正解率こそ大きく変化しないものの、再現率が著しく低下する。従って、単語数が少ない文書に対するフ

フィルタリングが困難であるという仮定は示されたといえる。

【0 0 6 7】

次に、同じ評価データに対し、URL階層構造を考慮したフィルタリングを行い、同様に正解率と再現率を求めた。この結果を表 7 に示す。

【0 0 6 8】

【表 7】

表 7 URL階層構造を考慮した自動
フィルタリングの正解率と再現率

min	正 解 率	再 現 率
5	91. 1%	68. 2%
10	91. 7%	79. 8%
15	90. 1%	80. 7%
20	90. 2%	82. 0%

この結果から、本発明による自動フィルタリング手法を取り入れることにより、高い正解率を維持したまま、再現率を大幅に増加させることができたことが明らかになった。これらの結果より、本発明の有効性が証明されたといえる。

【0 0 6 9】

次に、図 4 および図 5 を参照して、本発明の他の実施形態に係る自動フィルタリング装置について説明する。この実施形態の自動フィルタリング装置は、上述したように図 1～図 3 で説明した URL 階層構造を利用した情報自動フィルタリング装置に対して第三者判定によりフィルタリングを行う第三者判定フィルタリング処理部を付加するように構成したものであり、両フィルタリング処理を組み合わせることにより理想的なフィルタリングを達成しようとするものである。

【0 0 7 0】

図 4 に示す自動フィルタリング装置は、図 1～図 3 で説明した URL 階層構造を利用した情報自動フィルタリング装置 2 5 に対して第三者判定フィルタリング処理部 2 3 および該第三者判定フィルタリング処理部 2 3 で有害 URL を参照するために使用される有害 URL 一覧テーブル格納部 1 7 が付加されている。

【0071】

有害URL一覧テーブル格納部17は、有害情報を提供するURLを有害URL一覧テーブルとして格納しているものであり、第三者判定フィルタリング処理部23は、前記入力部1から入力されたHTML文書のURLを有害URL一覧テーブル格納部17の有害URL一覧テーブルに登録されている各URLと照合し、一致するURLがあるか否かを判定するものである。

【0072】

図5は、図4に示す自動フィルタリング装置の更に詳細な構成を示すブロック図である。図5に示す自動フィルタリング装置は、図1に示したURL階層構造を利用した情報自動フィルタリング装置を構成する入力部1、単語抽出部3、記憶部5、単語重みデータ格納部7、自動フィルタリング部9、有害上位ページ一覧テーブル格納部11、出力部13に加えて、図4の第三者判定フィルタリング処理部23に対応するURLリストに基づくフィルタリング部15および有害URL一覧テーブル格納部17を有している。

【0073】

このように構成される自動フィルタリング装置、すなわち第三者判定フィルタリング処理部によるURLリスト一覧とURL階層構造を利用した情報自動フィルタリング装置によるフィルタリング処理では、まずインターネット21を介して入力部1から入力されたHTML文書は、そのURLが有害URL一覧テーブル格納部17の有害URL一覧テーブルに登録されている各URLと照合され、一致するURLがあるか否かが判定される。そして、有害URL一覧テーブル格納部17の有害URL一覧テーブルに登録されたURLと一致する場合には、このURLが示す情報の提示は阻止される。

【0074】

URLリストに基づくフィルタリング部15による有害URL一覧テーブルを参照した判定の結果、有害URL一覧テーブル格納部17の有害URL一覧テーブルに登録されているURLと一致するものがない場合には、URL階層構造を利用した情報自動フィルタリング装置25によるフィルタリングが図1～図3で説明したように行われる。

【0075】

このように本実施形態では、第三者による判定に基づくフィルタリングとURL階層構造を利用したフィルタリングの両方が行われるため、有害情報を適確に検出して阻止することができる。

【0076】

【発明の効果】

以上説明したように、本発明によれば、HTML情報のURLが上位URLである場合、この上位URLが示す情報に対して自動フィルタリングを行い、その結果、該情報が不適切である場合、上位URLを不適切上位URL一覧に登録するとともに、該情報の提供を阻止し、上位URLでなかった場合、このURLを不適切上位URL一覧の各URLと照合し、一致するURLがある場合、このURLが示す情報の提示を阻止し、一致するものがない場合、該URLが示す情報に対して自動フィルタリングを行い、その結果、該情報が不適切である場合、該情報の提供を阻止するので、画像のみが提示されている少テキストページでもその不適切さを適確に判定して阻止することができ、正解率および再現率の両方を向上することができる。

【0077】

また、本発明によれば、URL階層構造を利用した情報自動フィルタリングに加えて、不適切な情報を提供するURLを不適切URL一覧として登録しておき、HTML情報のURLを不適切URL一覧の各URLと照合して、一致するURLがある場合、このURLが示す情報の提示を阻止する第三者判定に基づくフィルタリングを更に行うので、この第三者判定に基づくフィルタリングと上位URLを利用した自動フィルタリングの両方により更に完全にフィルタリングを行うことができる。

【図面の簡単な説明】

【図1】

本発明の一実施形態に係るURL階層構造を利用した情報自動フィルタリング装置の構成を示すブロック図である。

【図 2】

図 1 に示す URL 階層構造を利用した情報自動フィルタリング装置の作用を示すフローチャートである。

【図 3】

図 2 に示すフローチャートに使用されている単語重みの設定手順を示すフローチャートである。

【図 4】

本発明の他の実施形態に係る自動フィルタリング装置の概要構成を示す説明図である。

【図 5】

図 4 に示す自動フィルタリング装置の詳細な構成を示すブロック図である。

【図 6】

従来の自己判定に基づくフィルタリングを説明するための図である。

【図 7】

図 6 に示した自己判定に基づくフィルタリングの一例として RSACi と SafeSurf による評価結果の記述例を示す図である。

【図 8】

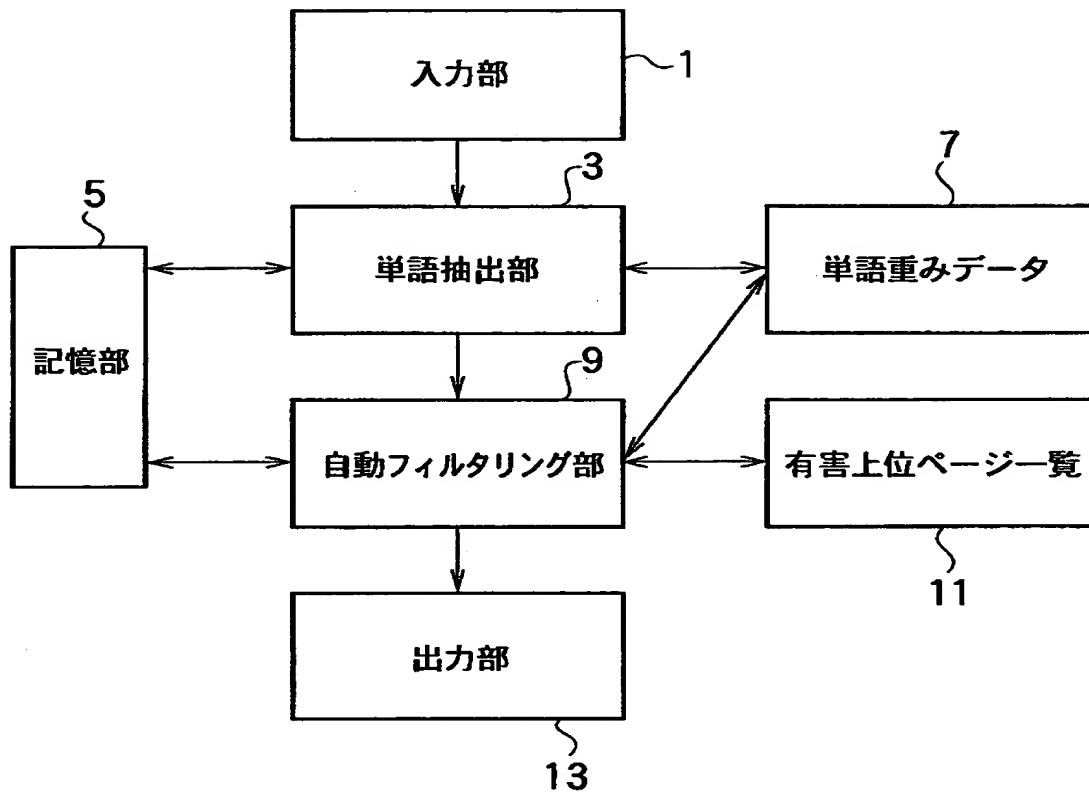
従来の第三者による判定に基づく有害情報フィルタリングを説明するための図である。

【符号の説明】

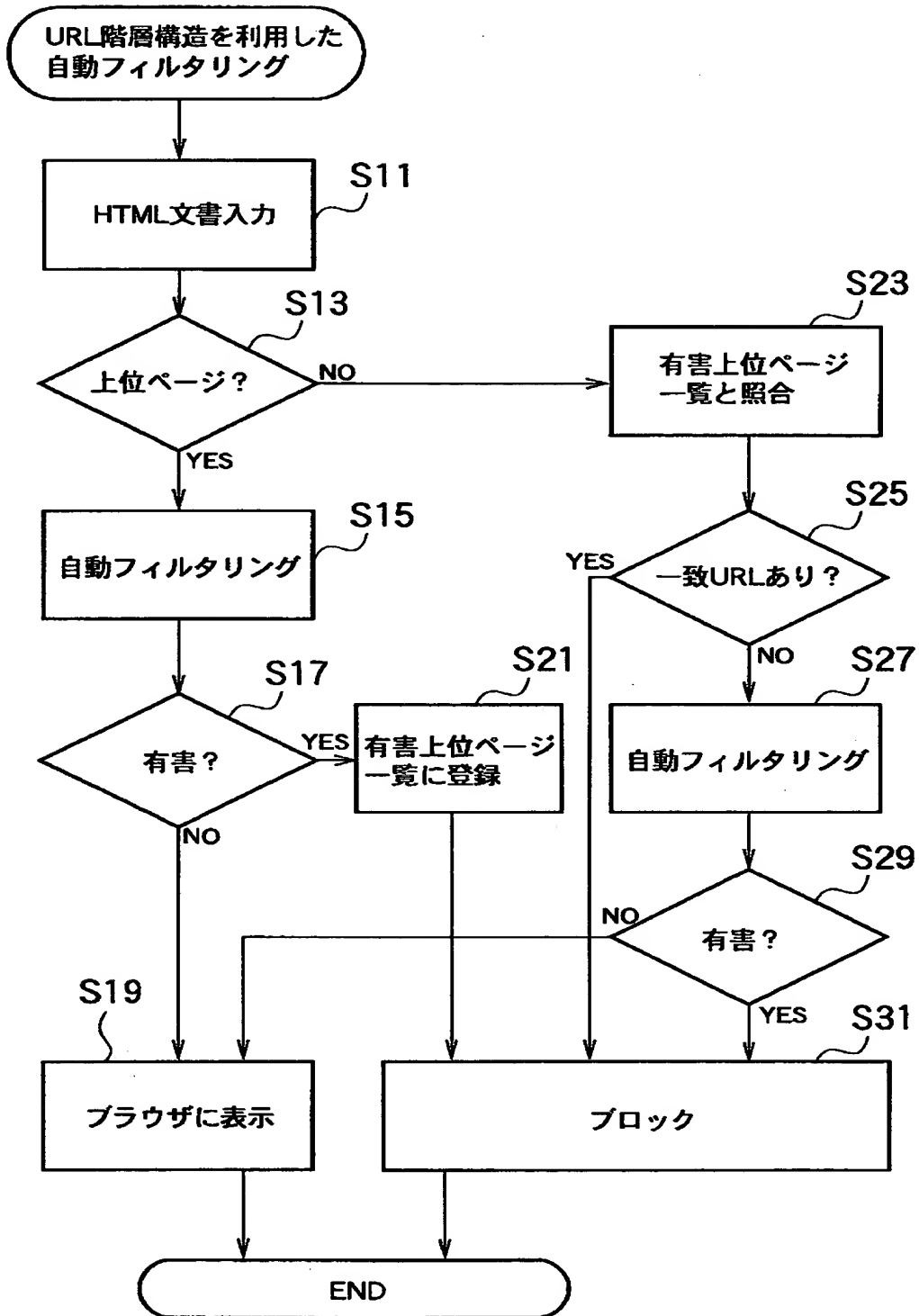
- 1 入力部
- 3 単語抽出部
- 7 単語重みデータ格納部
- 9 自動フィルタリング部
- 11 有害上位ページ一覧テーブル格納部
- 15 URL リストに基づくフィルタリング部
- 17 有害 URL 一覧テーブル格納部

【書類名】 図面

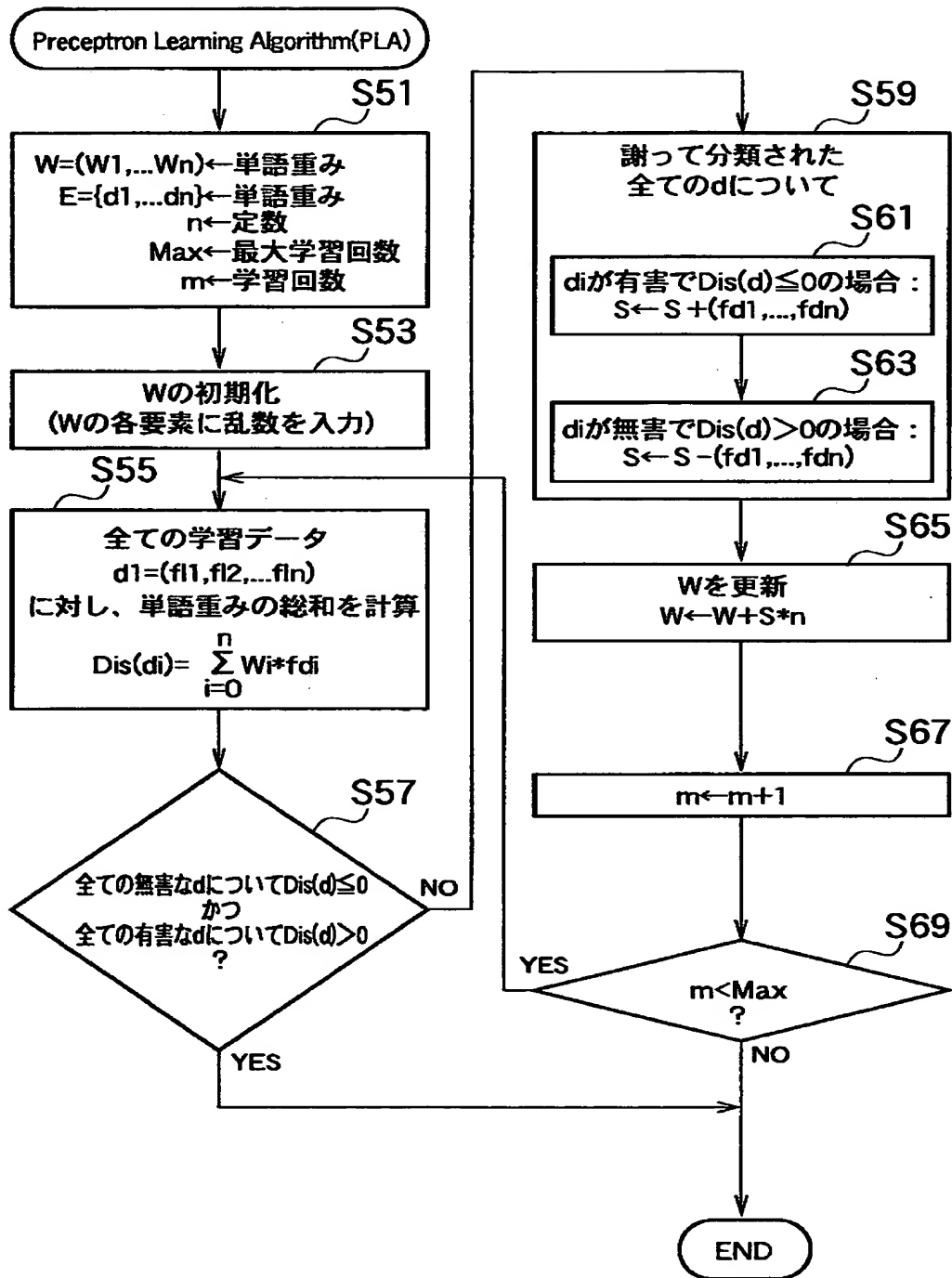
【図 1】



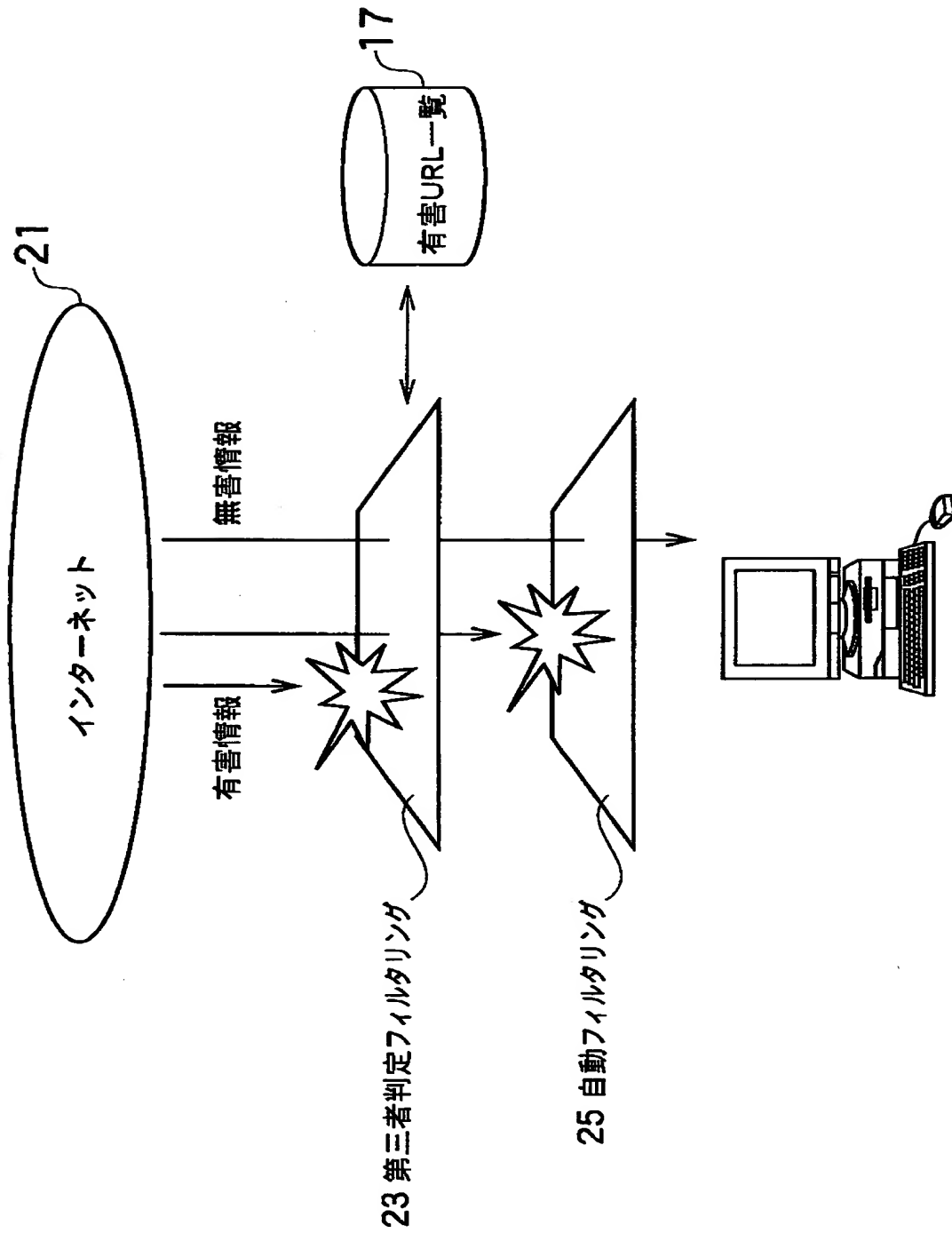
【図 2】



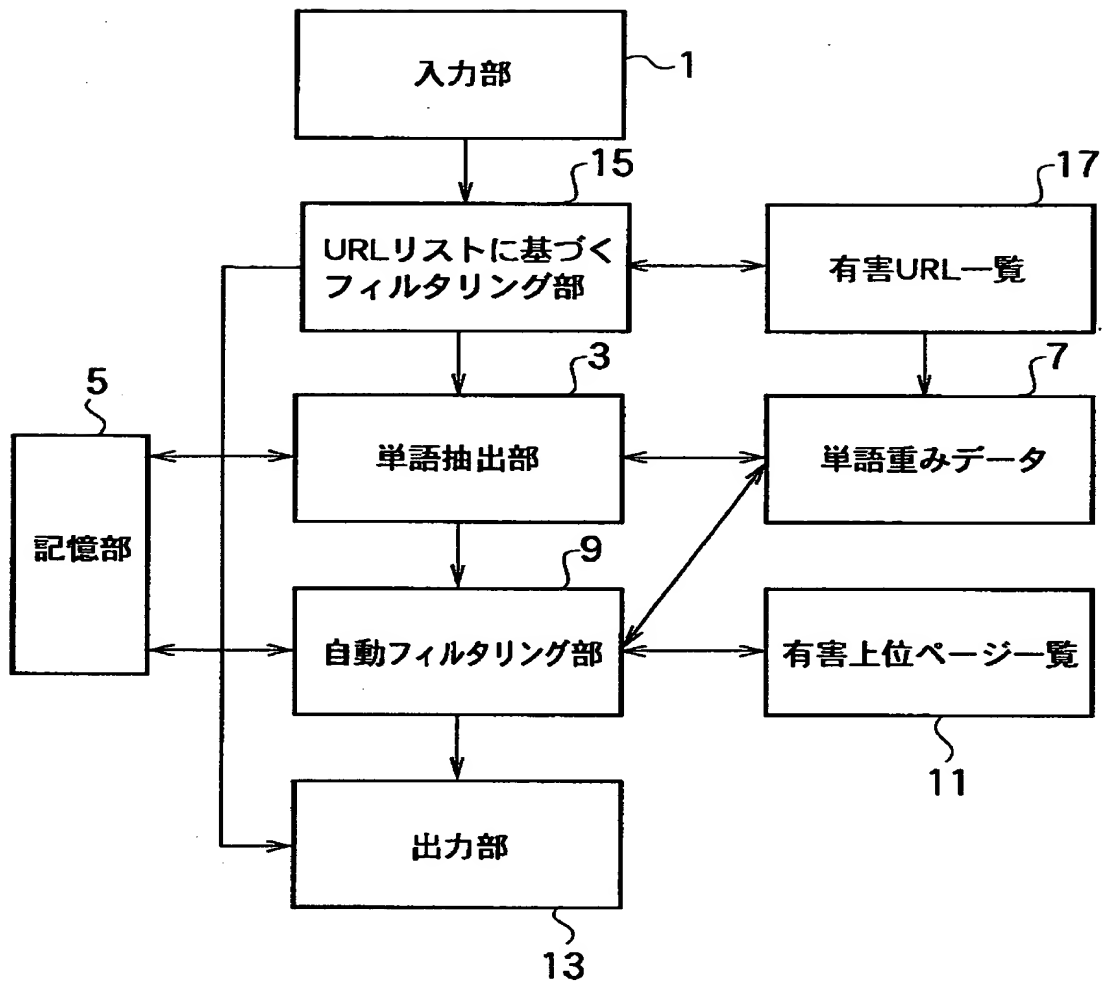
【図 3】



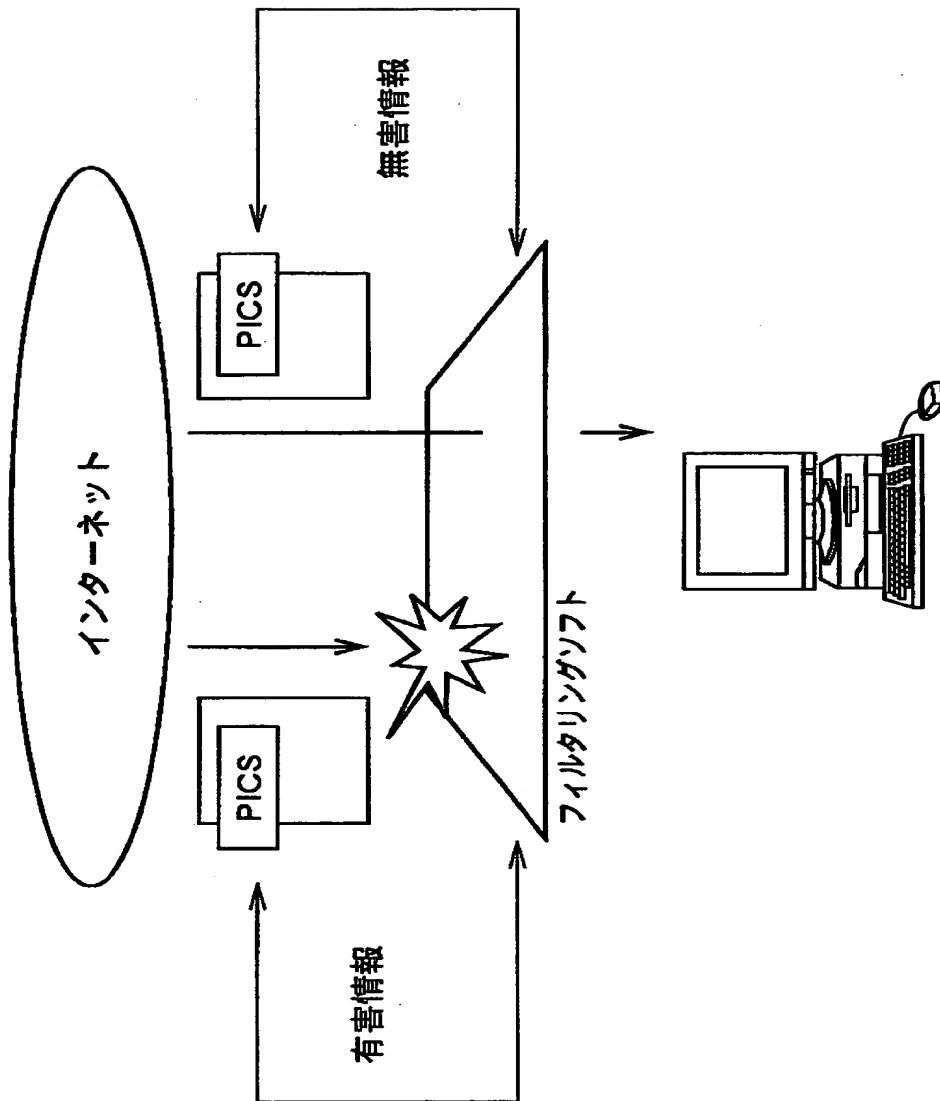
【図 4】



【図 5】



【図 6】



【図 7】

```

<HTML>
<HEAD>
<TITLE>
Example of RSAC1 and SafeSurf Rated Page
</TITLE>

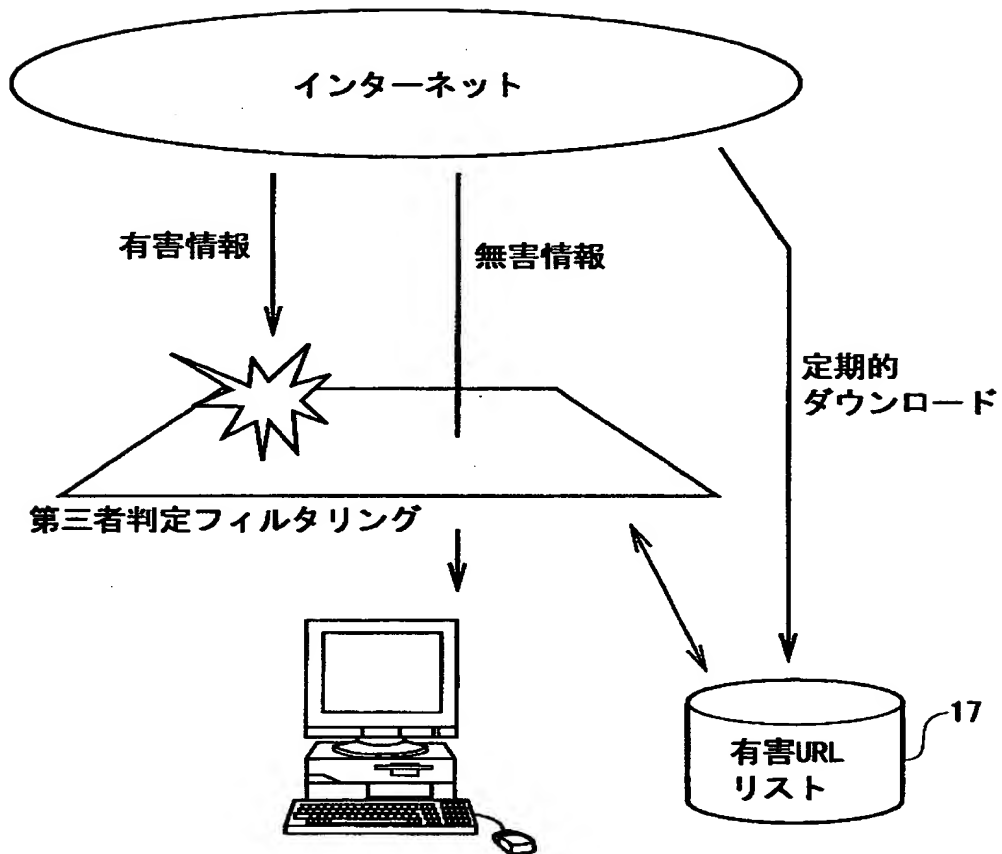
<META http-equiv="PICS-Label" content- '(PICS-1.0
"http://www.rsac.org/ratingav01.html"
1 gen true comment "RASC1 North America Serber" by
"marks@icrosys.com" for "http://www.samplesite.com" on
"1996.09.05T02:15-0800" exp "1997.09.05T02:15-0800"
r (n 2 s 0 v 1 1 2))'>

<META http-equiv="PICS-Label" content- '(PICS-1.1
"http://www.classify.org/safesurf/" 1 gen t for
"http://www.samplesite.com/"
r(SS~000 2 SS~100 1))'>

</HEAD>
<BODY>
...
```

RSACiとSafeSurfによる評価結果の記述例

【図 8】



【書類名】 要約書

【要約】

【課題】 階層構造になっているURLの上位URLを用いることにより正解率及び再現率共に向上し得るとともに、画像のみが掲載されている少テキストページに対しても内容の不適切さを適確に判定し得るURL階層構造を利用した情報自動フィルタリング方法および装置を提供する。

【解決手段】 入力されたHTML情報のURLが上位URLである場合、該上位URLが示す情報に対して自動フィルタリングを行い、該情報が不適切である場合、上位URLを不適切上位URL一覧に登録し、情報の提供を阻止し、上位URLでなかった場合、このURLを不適切上位URL一覧の各URLと照合し、一致するURLがある場合、情報の提示を阻止し、一致するものがない場合、該URLが示す情報に対して自動フィルタリングを行い、情報が不適切である場合、該情報の提供を阻止する。

【選択図】 図 2

出 願 人 履 歴 情 報

識別番号 [000001214]

1. 変更年月日	1998年12月 3日
[変更理由]	名称変更
住 所	東京都新宿区西新宿2丁目3番2号
氏 名	ケイディディ株式会社